RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: _10/516,338_
Source: _PYTO_
Date Processed by STIC: _12/9/04_

ENTERED

PCT

RAW SEQUENCE LISTING   DATE: 12/09/2004
PATENT APPLICATION: US/10/516,338   TIME: 14:36:59

Input Set : A:\6047252.app
Output Set: N:\CRF4\12092004\J516338.raw

```
 3 <110> APPLICANT: Astex Technology Limited
 4        Cosme, Jose
 5        Ward, Alison
 6        Vuillard, Laurent
 7        Williams, Pamela
 8        Hamilton, Bruce
10 <120> TITLE OF INVENTION: Methods of Purification of Cytochrome P450 Proteins
12 <130> FILE REFERENCE: AHBCP6047252
C--> 14 <140> CURRENT APPLICATION NUMBER: US/10/516,338
C--> 15 <141> CURRENT FILING DATE: 2004-11-30
17 <160> NUMBER OF SEQ ID NOS: 84
19 <170> SOFTWARE: PatentIn Ver. 2.1
21 <210> SEQ ID NO: 1
22 <211> LENGTH: 1428
23 <212> TYPE: DNA
24 <213> ORGANISM: Artificial Sequence
26 <220> FEATURE:
27 <223> OTHER INFORMATION: Description of Artificial Sequence: 2C19 (internal
28        deletion, and His tagged) coding sequence.
30 <400> SEQUENCE: 1
31 atggctaaga aaacgagctc taaagggcgg ccgcctggcc ccactcctct cccagtgatt 60
32 ggaaatatcc tacagataga tattaaggat gtcagcaaat ccttaaccaa tctctcaaaa 120
33 atctatggcc ctgtgttcac tctgtatttt ggcctggaac gcatggtggt gctgcatgga 180
34 tatgaagtgg tgaaggaagc cctgattgat cttggagagg agttttctgg aagaggccat 240
35 ttcccactgg ctgaaagagc taacagagga tttggaatcg ttttcagcaa tggaaagaga 300
36 tggaaggaga tccggcgttt ctccctcatg acgctgcgga attttgggat ggggaagagg 360
37 agcattgagg accgtgttca agaggaagcc cactgccttg tggaggagtt gagaaaaacc 420
38 aaggcttcac cctgtgatcc cactttcatc ctgggctgtg ctccctgcaa tgtgatctgc 480
39 tccattattt tccagaaacg tttcgattat aaagatcagc aatttcttaa cttgatggaa 540
40 aaattgaatg aaaacatcag gattgtaagc accccctgga tccagatatg caataatttt 600
41 cccactatca ttgattattt cccgggaacc cataacaaat tacttaaaaa ccttgctttt 660
42 atggaaagtg atattttgga gaaagtaaaa gaacaccaag aatcgatgga catcaacaac 720
43 cctcgggact ttattgattg cttcctgatc aaaatggaga aggaaaagca aaaccaacag 780
44 tctgaattca ctattgaaaa cttggtaatc actgcagctg acttacttgg agctgggaca 840
45 gagacaacaa gcacaaccct gagatatgct ctccttctcc tgctgaagca cccagaggtc 900
46 acagctaaag tccaggaaga gattgaacgt gtcgttggca gaaaccggag cccctgcatg 960
47 caggacaggg gccacatgcc ctacacagat gctgtggtgc acgaggtcca gagatacatc 1020
48 gacctcatcc ccaccagcct gccccatgca gtgacctgtg acgttaaatt cagaaactac 1080
49 ctcattccca agggcacaac catattaact tccctcactt ctgtgctaca tgacaacaaa 1140
50 gaatttccca acccagagat gtttgaccct cgtcactttc tgcatgaagg tggaaatttt 1200
51 aagaaaagta actacttcat gcctttctca gcaggaaaac ggatttgtgt gggagagggc 1260
52 ctggcccgca tggagctgtt tttattcctg accttcattt acagaactt taacctgaaa 1320
53 tctctgattg acccaaagga ccttgacaca actcctgttg tcaatggatt tgcttctgtc 1380
```

RAW SEQUENCE LISTING   DATE: 12/09/2004
PATENT APPLICATION: US/10/516,338   TIME: 14:36:59

Input Set : A:\6047252.app
Output Set: N:\CRF4\12092004\J516338.raw

```
54 ccgcccttct accagctctg cttcattcct gtccaccacc accactga                1428
57 <210> SEQ ID NO: 2
58 <211> LENGTH: 475
59 <212> TYPE: PRT
60 <213> ORGANISM: Artificial Sequence
62 <220> FEATURE:
63 <223> OTHER INFORMATION: Description of Artificial Sequence: Protein
64        sequence of 2C19 coded by SEQ ID NO: 1
66 <400> SEQUENCE: 2
67 Met Ala Lys Lys Thr Ser Ser Lys Gly Arg Pro Pro Gly Pro Thr Pro
68  1               5                  10                  15
70 Leu Pro Val Ile Gly Asn Ile Leu Gln Ile Asp Ile Lys Asp Val Ser
71              20                  25                  30
73 Lys Ser Leu Thr Asn Leu Ser Lys Ile Tyr Gly Pro Val Phe Thr Leu
74          35                  40                  45
76 Tyr Phe Gly Leu Glu Arg Met Val Val Leu His Gly Tyr Glu Val Val
77      50                  55                  60
79 Lys Glu Ala Leu Ile Asp Leu Gly Glu Glu Phe Ser Gly Arg Gly His
80  65                  70                  75                  80
82 Phe Pro Leu Ala Glu Arg Ala Asn Arg Gly Phe Gly Ile Val Phe Ser
83              85                  90                  95
85 Asn Gly Lys Arg Trp Lys Glu Ile Arg Arg Phe Ser Leu Met Thr Leu
86              100                 105                 110
88 Arg Asn Phe Gly Met Gly Lys Arg Ser Ile Glu Asp Arg Val Gln Glu
89          115                 120                 125
91 Glu Ala His Cys Leu Val Glu Glu Leu Arg Lys Thr Lys Ala Ser Pro
92      130                 135                 140
94 Cys Asp Pro Thr Phe Ile Leu Gly Cys Ala Pro Cys Asn Val Ile Cys
95 145                 150                 155                 160
97 Ser Ile Ile Phe Gln Lys Arg Phe Asp Tyr Lys Asp Gln Gln Phe Leu
98              165                 170                 175
100 Asn Leu Met Glu Lys Leu Asn Glu Asn Ile Arg Ile Val Ser Thr Pro
101             180                 185                 190
103 Trp Ile Gln Ile Cys Asn Asn Phe Pro Thr Ile Ile Asp Tyr Phe Pro
104         195                 200                 205
106 Gly Thr His Asn Lys Leu Leu Lys Asn Leu Ala Phe Met Glu Ser Asp
107     210                 215                 220
109 Ile Leu Glu Lys Val Lys Glu His Gln Glu Ser Met Asp Ile Asn Asn
110 225                 230                 235                 240
112 Pro Arg Asp Phe Ile Asp Cys Phe Leu Ile Lys Met Glu Lys Glu Lys
113             245                 250                 255
115 Gln Asn Gln Gln Ser Glu Phe Thr Ile Glu Asn Leu Val Ile Thr Ala
116         260                 265                 270
118 Ala Asp Leu Leu Gly Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu Arg
119     275                 280                 285
121 Tyr Ala Leu Leu Leu Leu Leu Lys His Pro Glu Val Thr Ala Lys Val
122     290                 295                 300
124 Gln Glu Glu Ile Glu Arg Val Val Gly Arg Asn Arg Ser Pro Cys Met
125 305                 310                 315                 320
```

```
127 Gln Asp Arg Gly His Met Pro Tyr Thr Asp Ala Val Val His Glu Val
128             325             330             335
130 Gln Arg Tyr Ile Asp Leu Ile Pro Thr Ser Leu Pro His Ala Val Thr
131             340             345             350
133 Cys Asp Val Lys Phe Arg Asn Tyr Leu Ile Pro Lys Gly Thr Thr Ile
134         355             360             365
136 Leu Thr Ser Leu Thr Ser Val Leu His Asp Asn Lys Glu Phe Pro Asn
137     370             375             380
139 Pro Glu Met Phe Asp Pro Arg His Phe Leu His Glu Gly Gly Asn Phe
140 385             390             395             400
142 Lys Lys Ser Asn Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Ile Cys
143             405             410             415
145 Val Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Phe Leu Thr Phe
146             420             425             430
148 Ile Leu Gln Asn Phe Asn Leu Lys Ser Leu Ile Asp Pro Lys Asp Leu
149         435             440             445
151 Asp Thr Thr Pro Val Val Asn Gly Phe Ala Ser Val Pro Pro Phe Tyr
152     450             455             460
154 Gln Leu Cys Phe Ile Pro Val His His His His
155 465             470             475
158 <210> SEQ ID NO: 3
159 <211> LENGTH: 1428
160 <212> TYPE: DNA
161 <213> ORGANISM: Artificial Sequence
163 <220> FEATURE:
164 <223> OTHER INFORMATION: Description of Artificial Sequence: 2C19 wild type
165          1B
167 <400> SEQUENCE: 3
168 atggctaaga aaacgagctc taaagggcgg ccgcctggcc ctactcctct cccagtgatt 60
169 ggaaatatcc tacagataga tattaaggat gtcagcaaat ccttaaccaa tctctcaaaa 120
170 atctatggcc ctgtgttcac tctgtatttt ggcctggaac gcatggtggt gctgcatgga 180
171 tatgaagtgg tgaaggaagc cctgattgat cttggagagg agttttctgg aagaggccat 240
172 ttcccactgg ctgaaagagc taacagagga tttggaatcg ttttcagcaa tggaaagaga 300
173 tggaaggaga tccggcgttt ctccctcatg acgctgcgga attttgggat ggggaagagg 360
174 agcattgagg accgtgttca agaggaagcc cgctgccttg tggaggagtt gagaaaaacc 420
175 aaagcttcac cctgtgatcc cactttcatc ctgggctgtg ctccctgcaa tgtgatctgc 480
176 tccattattt tccagaaacg tttcgattat aaagatcagc aatttcttaa cttgatggaa 540
177 aaattgaatg aaaacatcag gattgtaagc accccctgga tccagatatg caataatttt 600
178 cccactatca ttgattattt cccgggaacc cataacaaat tacttaaaaa ccttgctttt 660
179 atggaaagtg atattttgga gaaagtaaaa gaacaccaag aatcgatgga catcaacaac 720
180 cctcgggact ttattgattg cttcctgatc aaaatggaga aggaaaagca aaaccaacag 780
181 tctgaattca ctattgaaaa cttggtaatc actgcagctg acttacttgg agctgggaca 840
182 gagacaacaa gcacaaccct gagatatgct ctccttctcc tgctgaagca cccagaggtc 900
183 acagctaaag tccaggaaga gattgaacgt gtcgttggca gaaaccggag cccctgcatg 960
184 caggacaggg gccacatgcc ctacacagat gctgtggtgc acgaggtcca gagatacatc 1020
185 gacctcatcc ccaccagcct gccccatgca gtgacctgtg acgttaaatt cagaaactac 1080
186 ctcattccca agggcacaac catattaact tccctcactt ctgtgctaca tgacaacaaa 1140
187 gaatttccca acccagagat gtttgaccct cgtcactttc tggatgaagg tggaaatttt 1200
188 aagaaaagta actacttcat gcctttctca gcaggaaaac ggatttgtgt gggagagggc 1260
```

RAW SEQUENCE LISTING   DATE: 12/09/2004
PATENT APPLICATION: US/10/516,338   TIME: 14:36:59

Input Set : A:\6047252.app
Output Set: N:\CRF4\12092004\J516338.raw

189 ctggcccgca tggagctgtt tttattcctg accttcattt tacagaactt taacctgaaa 1320
190 tctctgattg acccaaagga ccttgacaca actcctgttg tcaatggatt tgcttctgtc 1380
191 ccgcccttct accagctctg cttcattcct gtccaccacc accactga   1428
194 <210> SEQ ID NO: 4
195 <211> LENGTH: 475
196 <212> TYPE: PRT
197 <213> ORGANISM: Artificial Sequence
199 <220> FEATURE:
200 <223> OTHER INFORMATION: Description of Artificial Sequence: Translation of
201   SEQ ID NO:3
203 <400> SEQUENCE: 4
204 Met Ala Lys Lys Thr Ser Ser Lys Gly Arg Pro Pro Gly Pro Thr Pro
205 1   5   10   15
207 Leu Pro Val Ile Gly Asn Ile Leu Gln Ile Asp Ile Lys Asp Val Ser
208   20   25   30
210 Lys Ser Leu Thr Asn Leu Ser Lys Ile Tyr Gly Pro Val Phe Thr Leu
211   35   40   45
213 Tyr Phe Gly Leu Glu Arg Met Val Val Leu His Gly Tyr Glu Val Val
214   50   55   60
216 Lys Glu Ala Leu Ile Asp Leu Gly Glu Glu Phe Ser Gly Arg Gly His
217 65   70   75   80
219 Phe Pro Leu Ala Glu Arg Ala Asn Arg Gly Phe Gly Ile Val Phe Ser
220   85   90   95
222 Asn Gly Lys Arg Trp Lys Glu Ile Arg Arg Phe Ser Leu Met Thr Leu
223   100   105   110
225 Arg Asn Phe Gly Met Gly Lys Arg Ser Ile Glu Asp Arg Val Gln Glu
226   115   120   125
228 Glu Ala Arg Cys Leu Val Glu Glu Leu Arg Lys Thr Lys Ala Ser Pro
229   130   135   140
231 Cys Asp Pro Thr Phe Ile Leu Gly Cys Ala Pro Cys Asn Val Ile Cys
232 145   150   155   160
234 Ser Ile Ile Phe Gln Lys Arg Phe Asp Tyr Lys Asp Gln Gln Phe Leu
235   165   170   175
237 Asn Leu Met Glu Lys Leu Asn Glu Asn Ile Arg Ile Val Ser Thr Pro
238   180   185   190
240 Trp Ile Gln Ile Cys Asn Asn Phe Pro Thr Ile Ile Asp Tyr Phe Pro
241   195   200   205
243 Gly Thr His Asn Lys Leu Leu Lys Asn Leu Ala Phe Met Glu Ser Asp
244   210   215   220
246 Ile Leu Glu Lys Val Lys Glu His Gln Glu Ser Met Asp Ile Asn Asn
247 225   230   235   240
249 Pro Arg Asp Phe Ile Asp Cys Phe Leu Ile Lys Met Glu Lys Glu Lys
250   245   250   255
252 Gln Asn Gln Gln Ser Glu Phe Thr Ile Glu Asn Leu Val Ile Thr Ala
253   260   265   270
255 Ala Asp Leu Leu Gly Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu Arg
256   275   280   285
258 Tyr Ala Leu Leu Leu Leu Leu Lys His Pro Glu Val Thr Ala Lys Val
259   290   295   300

```
261 Gln Glu Glu Ile Glu Arg Val Val Gly Arg Asn Arg Ser Pro Cys Met
262 305                     310                 315             320
264 Gln Asp Arg Gly His Met Pro Tyr Thr Asp Ala Val Val His Glu Val
265                 325                 330             335
267 Gln Arg Tyr Ile Asp Leu Ile Pro Thr Ser Leu Pro His Ala Val Thr
268                 340                 345             350
270 Cys Asp Val Lys Phe Arg Asn Tyr Leu Ile Pro Lys Gly Thr Thr Ile
271             355                 360             365
273 Leu Thr Ser Leu Thr Ser Val Leu His Asp Asn Lys Glu Phe Pro Asn
274         370             375             380
276 Pro Glu Met Phe Asp Pro Arg His Phe Leu Asp Glu Gly Gly Asn Phe
277 385             390             395             400
279 Lys Lys Ser Asn Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Ile Cys
280             405             410             415
282 Val Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Phe Leu Thr Phe
283         420             425             430
285 Ile Leu Gln Asn Phe Asn Leu Lys Ser Leu Ile Asp Pro Lys Asp Leu
286         435             440             445
288 Asp Thr Thr Pro Val Val Asn Gly Phe Ala Ser Val Pro Pro Phe Tyr
289     450             455             460
291 Gln Leu Cys Phe Ile Pro Val His His His His
292 465             470             475
295 <210> SEQ ID NO: 5
296 <211> LENGTH: 1443
297 <212> TYPE: DNA
298 <213> ORGANISM: Artificial Sequence
300 <220> FEATURE:
301 <223> OTHER INFORMATION: Description of Artificial Sequence: 2D6 encoding
302       nucleic acid
304 <400> SEQUENCE: 5
305 atggctaaaa aaacctcttc taaaggccga ccgccgggtc cgctgccgct gccaggcctg 60
306 ggtaacctgc tgcatgtgga cttccagaac accccgtact gcttcgacca gctgcgtcgt 120
307 cgtttcggtg acgtgttctc tctgcagctg gcttggaccc cggttgttgt tctgaacggt 180
308 ctggctgctg ttcgcgaagc tctggttacc cacggtgaag acaccgctga ccgtccgccg 240
309 gtcccgatca cccagatcct gggttttggt ccgcgttccc aaggtgtttt cctggctcgt 300
310 tacggaccgg cttggcgtga acagcgtcgt ttctctgttt ctaccctgcg taacctgggt 360
311 ctgggtaaaa aatctctgga acagtgggtt accgaagaag ctgcatgcct gtgcgctgct 420
312 ttcgctaacc actctggtcg tccgttccgt ccgaacggtc tgctggacaa agctgtttct 480
313 aacgttatcg cttctctgac ctgcggccgc cgtttcgaat acgacgaccc gcgtttcctg 540
314 cgtctgctgg acctggctca ggaaggtctg aaagaggagt ctggtttcct gcgtgaagtt 600
315 ctgaacgctg ttccggttct gctgcacatc ccagctctgg ctggtaaagt tctgcgtttc 660
316 cagaaagcat tcctgaccca gctggacgaa ctgctgaccg aacaccgtat gacctgggac 720
317 ccggctcagc cgccacgtga cctgaccgaa gctttcctgg ctgaaatgga aaaagctaaa 780
318 ggtaacccgg aatcttcttt caacgatgaa aatctgcgta tcgttgttgc tgacctgttc 840
319 tccgcgggta tggttaccac ctctaccacc ctggcttggg gtctgctgct gatgatcctg 900
320 cacccggatg tacagcgtcg tgttcagcag gaaatcgacg acgttattgg ccaggttcgt 960
321 cggccggaaa tgggtgacca ggctcacatg ccgtacacca ccgctgttat ccacgaagtt 1020
322 cagcgcttcg gtgacatcgt tccgctgggt atgacccaca tgacctctcg tgacatcgaa 1080
323 gttcagggtt ccgtatcccc gaaaggtacc accctgatca ccaacctgtc ttctgttctg 1140
```

VERIFICATION SUMMARY
PATENT APPLICATION: **US/10/516,338**

DATE: 12/09/2004
TIME: 14:37:00

Input Set : **A:\6047252.app**
Output Set: **N:\CRF4\12092004\J516338.raw**

L:14 M:270 C: Current Application Number differs, Replaced Current Application Number
L:15 M:271 C: Current Filing Date differs, Replaced Current Filing Date